An inexact scaled gradient projection method with applications in risk parity portfolios

Max Lemes
Universidade Federal de Goiás.
Instituto de Matemática e Estatística.
Goiânia, September 1, 2022.

Outline

The main problem

The main problem

We want to present an inexact version of the scaled gradient projection method for constrained convex optimization problem as follows

$$\min\{f(x) : \; x \in C\}, \tag{1}$$

where C is a closed and convex subset of $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function.

Scaled Gradient Projection Method[1]

Step 0. Choose $\sigma, \tau \in (0,1)$, $0 < \alpha_{\min} \leq \alpha_{\max}$. Let $x^0 \in C$ and set $k = 0$;

Step 1. Choose $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$ and a positive definite matrix D_k and take $w^k \in C$ as

$$w^k := \mathcal{P}_C^{D_k}(x^k - \alpha_k D_k^{-1} \nabla f(x^k))$$

If $w^k = x^k$, then **stop**; otherwise,

Step 2. Choose τ_k and define the next iterate x^{k+1} as

$$x^{k+1} = x^k + \tau_k (w^k - x^k). \tag{2}$$

and go back to the Step 1.

[1] S. Bonettini, R. Zanella, and L. Zanni. "A scaled gradient projection method for constrained image deblurring". In: *Inverse Problems* 25.1 (2009), pp. 015002, 23.

Scaled Gradient Projection Method

Let D be a $n \times n$ positive definite matrix and $\| \cdot \|_D : \mathbb{R}^n \to \mathbb{R}$ be the norm defined by

$$\|d\|_D := \sqrt{\langle Dd, d \rangle}, \quad \forall d \in \mathbb{R}^n.$$

For a fixed constant $\mu \geq 1$, *denote by $\mathcal{D}_\mu$ the set of symmetric positive definite matrices $n \times n$ with all eigenvalues contained in the interval $[\frac{1}{\mu}, \mu]$.*

- $\mathcal{D}_\mu$ is compact;
- If $D \in \mathcal{D}_\mu$, it follows that D^{-1} also belongs to $\mathcal{D}_\mu$;
- $\forall D \in \mathcal{D}_\mu$, we obtain

$$\frac{1}{\mu}\|d\|^2 \leq \|d\|_D^2 \leq \mu\|d\|^2, \qquad \forall d \in \mathbb{R}^n.$$

Exact Projection

Exact Projection

Definition (2.1)
The exact projection of the point $v \in \mathbb{R}^n$ onto C with respect to the norm $\|\cdot\|_D$, denoted by $\mathcal{P}_C^D(v)$, is defined by

$$\mathcal{P}_C^D(v) := \arg\min_{z \in C} \|z - v\|_D^2.$$

Lemma (2.2)
Let $v, w \in \mathbb{R}^n$. Then, $w = \mathcal{P}_C^D(v)$ if and only if $w \in C$ and

$$\langle D(v - w), y - w \rangle \leq 0,$$

for all $y \in C$.

Exact Projection

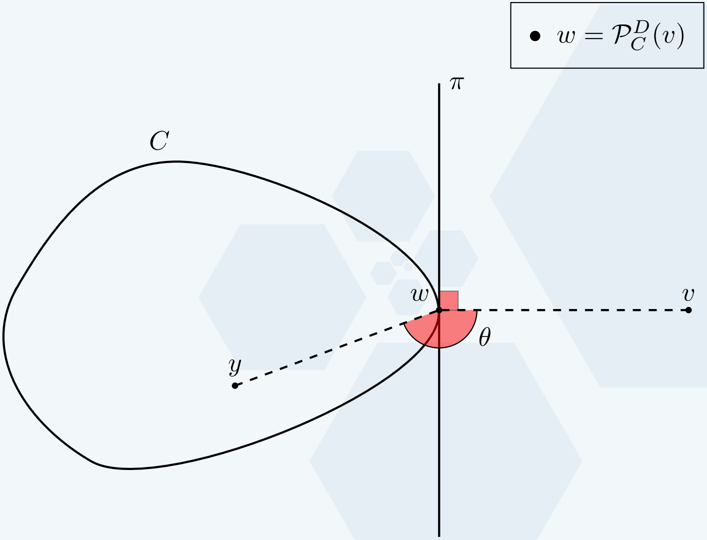


Figure 1: Exact projection of the point v onto C.

Inexact Projections

Inexact Projections

Definition (2.5[2])
The feasible inexact projection mapping, with respect to the norm $\|\cdot\|_D$, onto C relative to a point $u \in C$ and forcing parameter $\zeta \in (0,1]$, denoted by $\mathcal{P}^D_{C,\zeta}(u,\cdot) : \mathbb{R}^n \rightrightarrows C$, is the set-valued mapping defined as follows

$$\mathcal{P}^D_{C,\zeta}(u,v) := \left\{w \in C : \ \|w-v\|^2_D \leq \zeta \|\mathcal{P}^D_C(v) - v\|^2_D + (1-\zeta)\|u-v\|^2_D\right\}.$$

Each point $w \in \mathcal{P}^D_{C,\zeta}(u,v)$ is called a feasible inexact projection, with respect to the norm $\|\cdot\|_D$, of v onto C relative to u and forcing parameter $\zeta \in (0,1]$.

[2] Ernesto G. Birgin, José Mario Martínez, and Marcos Raydan. "Inexact spectral projected gradient methods on convex sets". In: *IMA J. Numer. Anal.* 23.4 (2003), pp. 539–559.

Inexact Projections

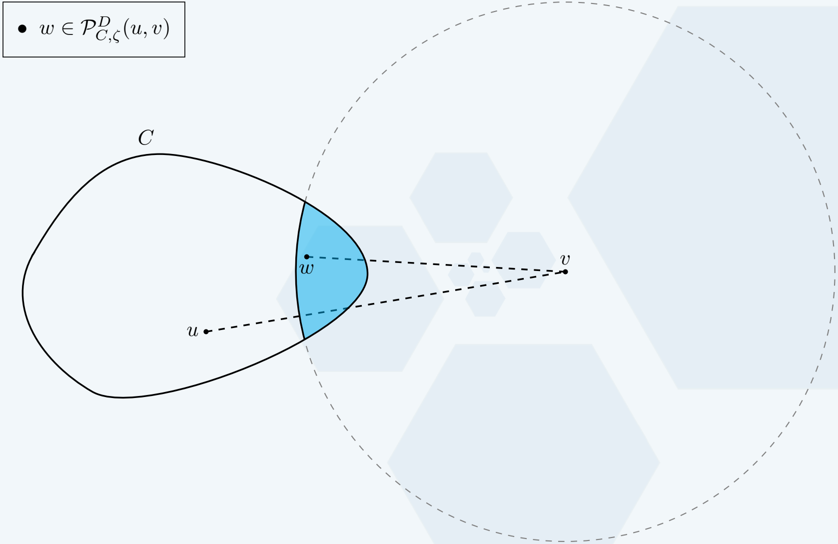


Figure 2: Feasible inexact projection of the point v onto C.

Inexact Projections

Definition (2.10[3][4])

The feasible inexact projection mapping, with respect to the norm $\|\cdot\|_D$, onto C relative to $u \in C$ and forcing parameter $\gamma \geq 0$, denoted by $\mathcal{R}^D_{C,\gamma}(u,\cdot) : \mathbb{R}^n \rightrightarrows C$, is the set-valued mapping defined as follows

$$\mathcal{R}^D_{C,\gamma}(u,v) := \left\{ w \in C : \ \langle D(v-w), y-w \rangle \leq \gamma \|w-u\|^2_D, \quad \forall\, y \in C \right\}.$$

Each point $w \in \mathcal{R}^D_{C,\gamma}(u,v)$ is called a feasible inexact projection, with respect to the norm $\|\cdot\|_D$, of v onto C relative to u and forcing parameter $\gamma \geq 0$.

---

[3] Saverio Salzo and Silvia Villa. "Inexact and accelerated proximal point algorithms". In: *J. Convex Anal.* 19.4 (2012), pp. 1167–1192.

[4] Fabiana R. de Oliveira, Orizon P. Ferreira, and Gilson N. Silva. "Newton's method with feasible inexact projections for solving constrained generalized equations". In: *Comput. Optim. Appl.* 72.1 (2019), pp. 159–177.

Inexact Projections

Let $u \in C$, $v \in \mathbb{R}^n$, $w \in \mathcal{R}^D_{C,\gamma}(u,v)$ be as stated in Definition 2.10, $0 \leq t < 1$ and

$$w_t = w + t(v - w).$$

Since

$$\langle D(v-w), y-w \rangle \leq \gamma \|w-u\|_D^2,$$

for all $y \in C$ we have

$$\langle D(v-w_t), y-w_t \rangle \leq (1-t)\left[\gamma\|w-u\|_D^2 - t\|v-w\|_D^2\right].$$

It follows that, for all

$$t \geq \frac{\gamma\|w-u\|_D^2}{\|v-w\|_D^2} \tag{3}$$

w_t satisfies

$$\langle D(v-w_t), y-w_t \rangle \leq 0. \tag{4}$$

Inexact Projections

Lemma (2.14)
Let $v \in \mathbb{R}^n$, $u \in C$, $\gamma \geq 0$ *and* $\zeta \in (0,1]$. *If* $0 \leq \gamma < 1/2$ *and* $\zeta = 1 - 2\gamma$, *then*

$$\mathcal{R}^D_{C,\gamma}(u,v) \subset \mathcal{P}^D_{C,\zeta}(u,v).$$

Proposition (2.17)
Let $v \in \mathbb{R}^n$, $u \in C$ *and assume that* C *is a bounded set. Then, for each* $0 < \gamma < 1/2$, *there exist* $0 < \zeta < 1$ *such that*

$$\mathcal{P}^D_{C,\zeta}(u,v) \subseteq \mathcal{R}^D_{C,\gamma}(u,v)$$

Inexact Projections

Lemma (2.18)
Let $x \in C$, $\alpha > 0$ *and* $z(\alpha) = x - \alpha D^{-1}\nabla f(x)$. *Take* $w(\alpha) \in \mathcal{P}^{D}_{C,\zeta}(x, z(\alpha))$ *with* $\zeta \in (0, 1]$. *Then, there hold*

(i) $\langle \nabla f(x), w(\alpha) - x\rangle \leq -\frac{1}{2\alpha}\|w(\alpha) - x\|^2_D + \frac{\zeta}{2\alpha}\left[\|\mathcal{P}^{D}_{C}(z(\alpha)) - z(\alpha)\|^2_D - \|x - z(\alpha)\|^2_D\right]$;

(ii) *the point* x *is stationary for problem* (1) *if, and only if,* $x \in \mathcal{P}^{D}_{C,\zeta}(x, z(\alpha))$;

(iii) *if* $x \in C$ *is a nonstationary point for problem* (1), *then* $\left\langle \nabla f(x), w(\alpha) - x \right\rangle < 0$.

Equivalently, if there exists $\bar{\alpha} > 0$ *such that* $\left\langle \nabla f(x), w(\bar{\alpha}) - x \right\rangle \geq 0$, *then* x *is stationary for problem* (1).

Inexact scaled gradient method

InexProj-SGM employing nonmonotone line search

Step 0. Choose $\sigma, \zeta_{\min} \in (0,1)$, $0 < \alpha_{\min} \leq \alpha_{\max}$ and $\mu \geq 1$. Let $x^0 \in C$, $\nu_0 \geq 0$ and set $k \leftarrow 0$.

Step 1. Choose positive real numbers α_k and ζ_k and a positive definite matrix D_k such that

$$\alpha_{\min} \leq \alpha_k \leq \alpha_{\max}, \qquad 0 < \zeta_{\min} < \zeta_k \leq 1, \qquad D_k \in \mathcal{D}_\mu.$$

Compute $w^k \in C$ as any feasible inexact projection with respect to the norm $\|\cdot\|_{D_k}$ of

$$z^k := x^k - \alpha_k D_k^{-1} \nabla f(x^k)$$

onto C relative to x^k with forcing parameter ζ_k, i.e.,

$$w^k \in \mathcal{P}_{C,\zeta_k}^{D_k}(x^k, z^k).$$

If $w^k = x^k$, then **stop** declaring convergence.

InexProj-SGM employing nonmonotone line search

Step 2.[5] Set $\tau_{\text{trial}} \leftarrow 1$. If

$$f\left(x^k + \tau_{\text{trial}}(w^k - x^k)\right) \le f(x^k) + \sigma\tau_{\text{trial}}\left\langle \nabla f(x^k), w^k - x^k \right\rangle + \nu_k, \tag{5}$$

then $\tau_k \leftarrow \tau_{\text{trial}}$, define the next iterate x^{k+1} as

$$x^{k+1} = x^k + \tau_k(w^k - x^k), \tag{6}$$

and go to **Step 3**. Otherwise, choose $\tau_{\text{new}} \in [\underline{\omega}\tau_{\text{trial}}, \bar{\omega}\tau_{\text{trial}}]$, set $\tau_{\text{trial}} \leftarrow \tau_{\text{new}}$, and repeat test (5).

Step 3.[6] Take $\delta_{k+1} \in [\delta_{\min}, 1]$ and choose $\nu_{k+1} \in \mathbb{R}$ satisfying

$$0 \le \nu_{k+1} \le (1 - \delta_{k+1})\left[f(x^k) + \nu_k - f(x^{k+1})\right].$$

Set $k \leftarrow k + 1$ and go to **Step 1**.

[5] Ernesto G. Birgin, José Mario Martínez, and Marcos Raydan. "Inexact spectral projected gradient methods on convex sets". In: *IMA J. Numer. Anal.* 23.4 (2003), pp. 539–559.

[6] Geovani N. Grapiglia and Ekkehard W. Sachs. "On the worst-case evaluation complexity of non-monotone line search algorithms". In: *Comput. Optim. Appl.* 68.3 (2017), pp. 555–577.

Nonmonotone line search

Remarks

There are several ways of choosing ν_k

(i) If $\nu_k = 0$ and $\delta_{k+1} = 1$, the line search (6) is the well-known Armijo line search.

(ii) Let $M > 0$ be an integer parameter and $f(x^{\ell(k)}) := \max\{f(x^{k-j}) \,|\, 0 \leq j \leq \min\{k, M\}\}$. If

$$\nu_k = f(x^{\ell(k)}) - f(x^k) \text{ and }$$
$$\delta_{k+1} \leq \frac{f(x^{\ell(k)}) - f(x^{\ell(k+1)})}{f(x^{\ell(k)}) - f(x^{k+1})} \tag{7}$$

the line search (6) is the same defined by Grippo, Lampariello and Lucidi[7].

[7] L. Grippo, F. Lampariello, and S. Lucidi. "A nonmonotone line search technique for Newton's method". In: *SIAM J. Numer. Anal.* 23.4 (1986), pp. 707–716.

Nonmonotone line search

(iii) Let $0 \leq \eta_{min} \leq \eta_{max} < 1$, $c_0 = f(x_0)$ and $q_0 = 1$. Choose $\eta_k \in [\eta_{min}, \eta_{max}]$ and set

$$q_{k+1} = \eta_k q_k + 1, \qquad c_{k+1} = (\eta_k q_k c_k + f(x^{k+1}))/q_{k+1}, \qquad \forall k \in \mathbb{N}.$$

If

$$\begin{aligned} \nu_k &= c_k - f(x^k) \text{ and} \\ \delta_{k+1} &= \frac{1}{q_{k+1}} \end{aligned} \tag{8}$$

the line search (6) is the same defined by Zhang and Hager[8].

[8] H. Zhang and W. W. Hager. "A nonmonotone line search technique and its application to unconstrained optimization". In: *SIAM J. Optim.* 14.4 (2004), pp. 1043–1056.

Partial asymptotic convergence

Partial asymptotic convergence analysis

Lemma (3.2)
*As consequence of **Step 3** the sequence* $\big(f(x^k) + \nu_k\big)_{k\in\mathbb{N}}$ *is non-increasing.*

Proposition (3.3)
Assume that $\lim\limits_{k\to+\infty} \nu_k = 0$*. Then, Algorithm InexProj-SGM stops in a finite number of iterations at a stationary point of problem* (1)*, or generates an infinite sequence* $(x^k)_{k\in\mathbb{N}}$ *for which every cluster point is stationary for problem* (1)*.*

Partial asymptotic convergence analysis

Corollary (3.4)

If $\delta_{min} > 0$, *then* $\sum_{k=0}^{+\infty} \nu_k < +\infty$. *Consequently,* $\lim_{k \to +\infty} \nu_k = 0$.[9]

Remarks

- Armijo line search and Zhang, Hager nonmonotone line search strategy satisfies a condition $\delta_{min} > 0$.
- For the Grippo, Lampariello and Lucidi nonmonotone line search strategy, we can only guarantee that $\delta_{min} \geq 0$.

Hence, we need deal with this case separately.

[9] Geovani N. Grapiglia and Ekkehard W. Sachs. "On the worst-case evaluation complexity of non-monotone line search algorithms". In: *Comput. Optim. Appl.* 68.3 (2017), pp. 555–577.

Partial asymptotic convergence analysis

Proposition (3.5)

Assume that the sequence $(x^k)_{k\in\mathbb{N}}$ is generated by Algorithm InexProj-SGM with the nonmonotone line search (7), i.e., $\nu_k = f(x^{\ell(k)}) - f(x^k)$ for all $k \in \mathbb{N}$. In addition, assume that the level set $C_0 := \{x \in C : \ f(x) \leq f(x^0)\}$ is bounded and $\nu_0 = 0$. Then, $\lim_{k\to+\infty} \nu_k = 0$.

Remark

- Combining Propositions 3.3 and 3.5, we conclude that $(x^k)_{k\in\mathbb{N}}$ is either finite terminating at a stationary point of problem (1), or infinite, and every cluster point of $(x^k)_{k\in\mathbb{N}}$ is stationary for problem (1).

Therefore, we have an alternative proof for the result obtained in [4, Theorem 2.1][10].

[10] Ernesto G. Birgin, José Mario Martínez, and Marcos Raydan. "Inexact spectral projected gradient methods on convex sets". In: *IMA J. Numer. Anal.* 23.4 (2003), pp. 539–559.

Full asymptotic convergence

Full asymptotic convergence analysis

We will prove, under suitable assumptions, the full convergence of the sequence $(x^k)_{k\in\mathbb{N}}$. For this end, we assume that in $\mathbf{Step\,1}$ of Algorithm InexProj-SGM:

A1. For all $k \in \mathbb{N}$, we take $w^k \in \mathcal{R}^{D_k}_{C,\gamma_k}(x^k, z^k)$ with $\gamma_k = (1-\zeta_k)/2$.

A2. For all $k \in \mathbb{N}$, we take $0 \leq \nu_k$ such that $\sum_{k=0}^{+\infty} \nu_k < +\infty$.

To prove the full convergence of the sequence $(x^k)_{k\in\mathbb{N}}$ satisfying **A1** and **A2** we consider an additional assumption on the sequence $(D_k)_{k\in\mathbb{N}} \subset \mathcal{D}_\mu$ as follows.

A3. For all $k \in \mathbb{N}$, $(1+\eta_k)D_k - D_{k+1}$ is a positive semidefinite matrix, for some sequence $(\eta_k)_{k\in\mathbb{N}} \subset [0, +\infty)$ such that $\sum_{k\in\mathbb{N}} \eta_k < \infty$.

Armijo line search and nonmonotone line search strategies defined by (8) satisfies **A2**.

Full asymptotic convergence analysis

Theorem (3.9)
If f is a convex function and $(x^k)_{k\in\mathbb{N}}$ has no cluster points, then $\Omega^* = \varnothing$, $\lim_{k\to\infty} \|x^k\| = +\infty$, and $\inf_{k\in\mathbb{N}} f(x^k) = \inf\{f(x) : x \in C\}$.

Corollary (3.10)
If f is a convex function and $(x^k)_{k\in\mathbb{N}}$ has at least one cluster point, then $(x^k)_{k\in\mathbb{N}}$ converges to a stationary point of problem (1).

Theorem (3.11)
Assume that f is a convex function and $\Omega^* \neq \varnothing$. Then, $(x^k)_{k\in\mathbb{N}}$ converge to an optimal solution of problem (1).

Iteration-complexity bound

Interation-complexity bound

Besides assuming that in $\textbf{Step 1}$ of Algorithm InexProj-SGM we take $(x^k)_{k\in\mathbb{N}}$ satisfying **A1** and **A2**, we also need the following assumption.

A3. The gradient ∇f of f is Lipschitz continuous with constant $L > 0$.

Lemma (3.12)
The steepsize τ_k in Algorithm InexProj-SGM satisfies $\tau_k \geq \tau_{\min}$,

where

$$\tau_{\min} := \min\left\{1, \frac{\tau(1-\sigma)}{\alpha_{\max}\mu L}\right\}.$$

Interation-complexity bound

Theorem (3.13)
For every $N \in \mathbb{N}$, the following inequality holds

$$\min\left\{\|w^k - x^k\| : \ k = 0, 1 \ldots, N-1\right\} \leq \sqrt{\frac{2\alpha_{\max}\mu\left(f(x^0) - f^* + \sum_{k=0}^{\infty}\nu_k\right)}{\sigma\tau_{\min}}}\frac{1}{\sqrt{N}}.$$

Theorem (3.17)
Let f be a convex function on C. Then, for every $N \in \mathbb{N}$, there holds

$$\min\left\{f(x^k) - f^* : \ k = 0, 1 \ldots, N-1\right\} \leq \frac{\|x^0 - x^*\|_{D_0}^2 + \xi\left[f(x^0) - f^* + \sum_{k=0}^{\infty}\nu_k\right]}{2\alpha_{\min}\tau_{\min}}\frac{1}{N}.$$

Interation-complexity bound

Theorem (3.15)
For a given $\epsilon > 0$, the number of function evaluations in Algorithm InexProj-SGM are at most

$$1 + \left(\frac{2\alpha_{\max}\mu \left(f(x^0) - f^* + \sum_{k=0}^{\infty} \nu_k \right)}{\sigma \tau_{\min}} \frac{1}{\epsilon^2} + 1 \right) \left(\frac{\log(\tau_{\min})}{\log(\tau)} + 1 \right),$$

to compute x^k and w^k such that $\|w^k - x^k\| \leq \epsilon$.

Theorem (3.16)
Let f be a convex function on C. For a given $\epsilon > 0$, the number of function evaluations in Algorithm InexProj-SGM are at most

$$1 + \left(\frac{\|x^0 - x^*\|_{D_0}^2 + \xi \left(f(x^0) - f^* + \sum_{k=0}^{\infty} \nu_k \right)}{2\alpha_{\min}\tau_{\min}} \frac{1}{\epsilon} + 1 \right) \left(\frac{\log(\tau_{\min})}{\log(\tau)} + 1 \right),$$

to compute x^k such that $f(x^k) - f^* \leq \epsilon$.

Numerical experiments

Numerical experiments

Given A and B two $m \times n$ matrices, with $m \geq n$, and $c \in \mathbb{R}$, we consider the matrix function $f : \mathbb{R}^{n\times n} \to \mathbb{R}$ given by:

$$f(X) := \frac{1}{2}\|AX - B\|_F^2 + \sum_{i=1}^{n-1}\left[c\left(X_{i+1,i+1} - X_{i,i}^2\right)^2 + (1 - X_{i,i})^2\right],$$

which combines a least squares term with a Rosenbrock-type function. $X_{i,j}$ stands for the ij-element of the matrix X and $\|\cdot\|_F$ denotes the Frobenius matrix norm, i.e., $\|A\|_F := \sqrt{\langle A, A\rangle}$ where the inner product is given by $\langle A, B\rangle = \mathrm{tr}(A^T B)$.

Numerical experiments

Problem I[11]:

$$\begin{array}{ll} \min & f(X) \\ \text{s.t.} & X \in SDD^+, \\ & L \le X \le U, \end{array}$$

where SDD^+ is the cone of symmetric and diagonally dominant real matrices with positive diagonal, i.e.,

$$SDD^+ := \{X \in \mathbb{R}^{n\times n} \mid X = X^T,\ X_{i,i} \ge \sum_{j\neq i} |X_{i,j}|\ \forall i\}, \tag{9}$$

L and U are given $n \times n$ matrices, and $L \le X \le U$ means that $L_{i,j} \le X_{i,j} \le U_{i,j}$ for all i, j.

For Problem I, we used the Dykstra's algorithm described in [4], see also [15] to compute the inexact projection. In this case, $SDD^+ = \cap_{i=1}^{n} SDD_i^+$, where

$$SDD_i^+ := \{X \in \mathbb{R}^{n\times n} \mid X = X^T,\ X_{i,i} \ge \sum_{j\neq i} |X_{i,j}|\} \text{ for all } i = 1, \ldots, n. \tag{10}$$

[11]Ernesto G. Birgin, José Mario Martínez, and Marcos Raydan. "Inexact spectral projected gradient methods on convex sets". In: *IMA J. Numer. Anal.* 23.4 (2003), pp. 539–559.

Numerical experiments

Problem II[12][13]**:**

$$\begin{array}{ll} \min & f(X) \\ \text{s.t.} & X \in \mathbb{S}^n_+, \\ & \operatorname{tr}(X) = 1, \end{array}$$

where $\mathbb{S}^n_+$ is the cone of symmetric and positive semidefinite real matrices. The feasible set of Problem II was known as *spectrahedron* and appears in several interesting applications.

We considered a variant of the Frank-Wolfe algorithm proposed in [2], which improves the convergence rate and the total time complexity of the classical Frank-Wolfe method. This algorithm specialized for the projection problem over the spectrahedron is carefully described in [1].

[12] Zeyuan Allen-Zhu et al. "Linear convergence of a Frank-Wolfe type algorithm over trace-norm balls". In: *Advances in Neural Information Processing Systems*. 2017, pp. 6191–6200.

[13] D.S. Gonçalves, M.A. Gomes-Ruggiero, and C. Lavor. "A projected gradient method for optimization over density matrices". In: *Optimization Methods and Software* 31.2 (2016), pp. 328–341.

Numerical experiments

We are interested in the spectral gradient version of the SPG method, so we set $D_k := I$ for all k, $\alpha_0 := \min(\alpha_{\max}, \max(\alpha_{\min}, 1/\|\nabla f(x^0)\|))$ and, for $k > 0$,

$$\alpha_k := \begin{cases} \min(\alpha_{\max}, \max(\alpha_{\min}, \langle s^k, s^k \rangle / \langle s^k, y^k \rangle)), & \text{if } \langle s^k, y^k \rangle > 0 \\ \alpha_{\max}, & \text{otherwise,} \end{cases} \tag{11}$$

where $s^k := X^k - X^{k-1}$, $y^k := \nabla f(X^k) - \nabla f(X^{k-1})$, $\alpha_{\min} = 10^{-10}$, and $\alpha_{\max} = 10^{10}$.

Concerning the stopping criterion, all runs were stopped at an iterate X^k declaring convergence if

$$\max_{i,j}(|X^k_{i,j} - W^k_{i,j}|) \leq 10^{-6},$$

where $W^k \in \mathcal{P}^{D_k}_{C,\zeta_k}(x^k, z^k)$.

Influence of the inexact projection

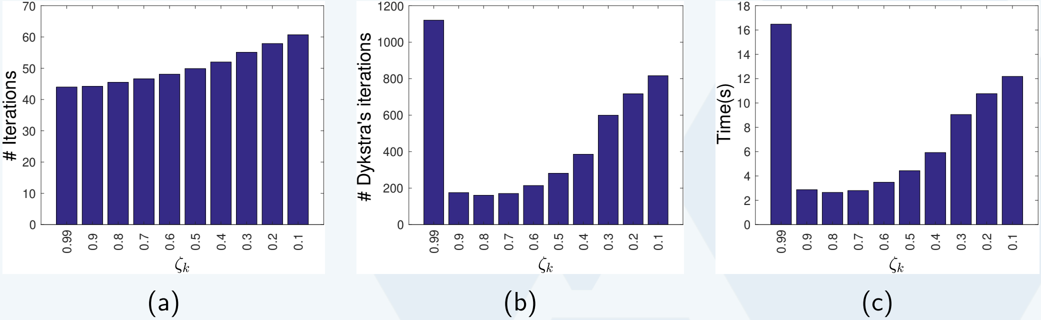


Figure 3: Results for 10 instances of Problem I using $n = 100$, $m = 200$, and $c = 10$. Average number of: (a) iterations; (b) Dykstra's iterations; (c) CPU time in seconds needed to reach the solution for different choices of ζ_k.

Influence of the inexact projection

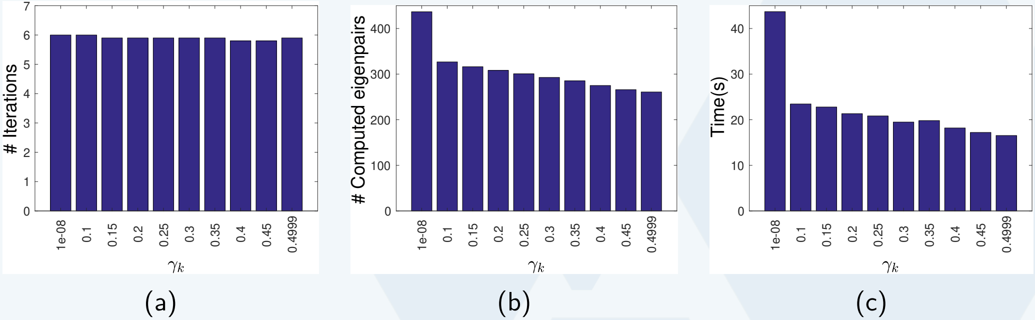
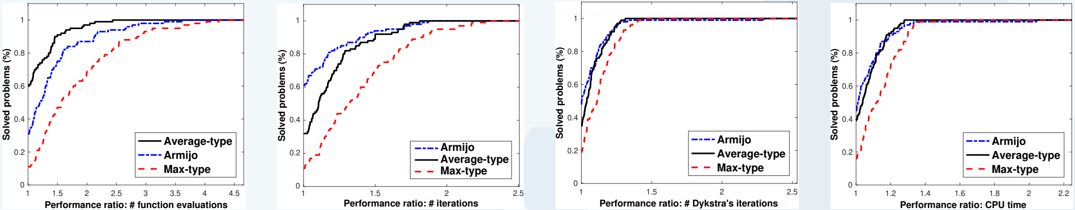


Figure 4: Results for 10 instances of Problem II using $n = 800$, $m = 1000$, and $c = 100$. Average number of: (a) iterations; (b) computed eigenpairs; (c) CPU time in seconds needed to reach the solution for different choices of γ_k.

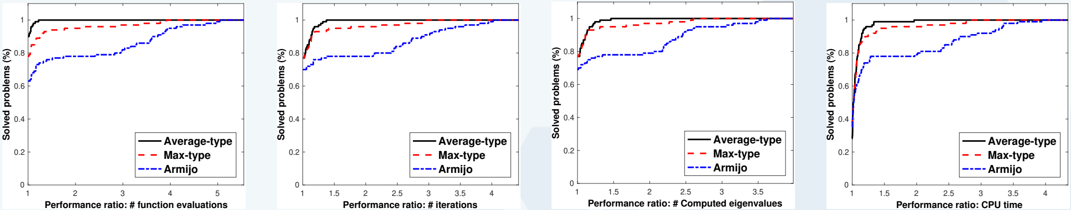
Influence of the line search scheme



(a) (b) (c) (d)

Figure 5: Performance profiles for Problem I considering the SPG method with the Armijo, the Average-type, and the Max-type line searches strategies using as performance measurement: (a) number of function evaluations; (b) number of (outer) iterations; (c) number of Dykstra's iterations; (d) CPU time.

Influence of the line search scheme



(a) (b) (c) (d)

Figure 6: Performance profiles for Problem II considering the SPG method with the Armijo, the Average-type, and the Max-type line searches strategies using as performance measurement: (a) number of function evaluations; (b) number of (outer) iterations; (c) number of computed eigenpairs; (d) CPU time.

Risk Parity Portfolios

Risk Measures

Definition (5.1)
A portfolio with n assets, $A_1, A_2, \dots, A_n$, is a vector $x^\top = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, each coordinate x_i being the weight of the capital invested in asset A_i.

One way to measure the risk of portfolio is by its volatility, i.e, the standard deviation of its daily returns.

If we denote by μ and Σ the vector of expected returns and the covariance matrix of asset returns respectively, we deduce that the expected return of portfolio x equals:

$$\mu(x) = x^\top \mu. \tag{12}$$

Its variance is given by:

$$\sigma^2(x) = x^\top \Sigma x. \tag{13}$$

Minimum Variance Portfolio (MVP)

The minimum variance portfolio (MVP) of Markowitz[14], consists of minimize variance with fully invested capital, achieved by solving the problem:

$$\begin{aligned} &\min_{x} \; \{x^\top \Sigma x\}, \\ \text{s.t.} \; &\left\{ \mathbf{1}^\top x = 1. \right. \end{aligned} \tag{14}$$

where $\mathbf{1}^\top = (1, 1, \ldots, 1)$.

[14] Harry Markowitz. "Portfolio Selection". In: *The Journal of Finance* 7.1 (1952), pp. 77–91.

Risk Parity Portfolio (RPP)

Let $x^\top = (x_1, x_2, \ldots, x_n)$ be a portfolio with n assets and $\mathcal{R}(x)$ be a differentiable, homogeneous risk measure of x. We have

$$\mathcal{R}(x) = \frac{d}{d\lambda}\mathcal{R}(\lambda x) = \sum_{i=1}^{n} x_i \frac{\partial \mathcal{R}(x)}{\partial x_i}. \tag{15}$$

We define the risk contribution of asset i as

$$\mathcal{RC}_i(x) = x_i \frac{\partial \mathcal{R}(x)}{\partial x_i}. \tag{16}$$

Therefore, the risk may be written as follows

$$\mathcal{R}(x) = \sum_{i=1}^{n} \mathcal{RC}_i(x) \tag{17}$$

which is known as **Euler's Allocation Principle**.

Risk Parity Portfolio (RPP)

It may be interesting for the investor to choose different risk levels for different assets. In this case the investor choose the percentage of risk, b_i, that each asset should have in the portfolio, that is,

$$\mathcal{RC}_i(x) = b_i \mathcal{R}(x), \tag{18}$$

$b_i \geq 0$ for all i, and $\mathbf{1}^\top b = 1$, with $b^\top = (b_1, b_2, \ldots, b_n)$. A portfolio distribution based on equation (18) is called a *risk budgeting portfolio* (RBP). When $b_i = 1/n$, for all i, the distribution is called *risk parity portfolio* (RPP) or *equal risk portfolio* (ERP).

In general, find the risk budgeting portfolio consists in solving the following non-linear system

$$\begin{aligned} \mathcal{RC}_i(x) &= b_i \mathcal{R}(x), \\ \text{s.t.} \quad &\begin{cases} b_i \geq 0, \\ x_i \geq 0, \\ \mathbf{1}^\top b = 1, \\ \mathbf{1}^\top x = 1. \end{cases} \end{aligned} \tag{19}$$

Risk Parity Portfolio(RPP)

Kaya and Lee[15] show that, in the Gaussian case, the risk budgeting portfolio may be found solving the optimization problem

$$\min_{x>\mathbf{0}} \{-b^\top \ln(x)\};$$
$$\text{s.t.} \quad \begin{cases} \sigma^2(x) \leq \sigma_0, \\ \mathbf{1}^\top x = 1, \end{cases} \tag{20}$$

σ_0 being a volatility target.

[15] H. Kaya and W. Lee. "Demystifying risk parity". In: *Neuberger Bermann* (2012).

Backtest Study

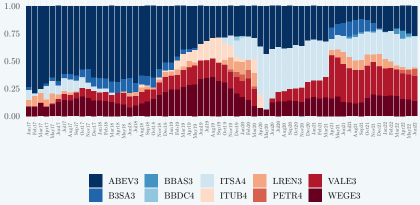
We present an application for Algorithm InexProj-SGM, a backtest study of a portfolio build with the 10 biggest companies on the brazilian market:

Ambev S.A. (ABEV3), B3 S.A. (B3SA3), Banco do Brasil S.A. (BBAS3), Banco Bradesco S.A. (BBDC4), Itaúsa S.A. (ITSA3), Itaú Unibanco S.A. (ITUB4), Lojas Renner S.A. (LREN3), Petróleo Brasileiro S.A (PETR4), Vale S.A. (VALE3) and WEG S.A. (WEGE3).

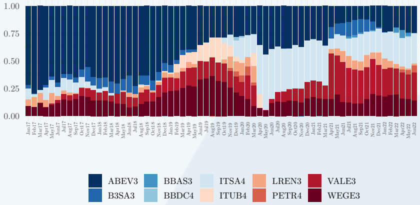
We did a backtest, building 2 portfolios with monthly rebalance:

- One was the minimum variance portfolio (MVP),
- the other the risk parity portfolio (RPP).

Data pertaining to the period between January 2016 and June 2022, was taken from Yahoo Finance (`http://finance.yahoo.com`). For rebalancing, the volatilities were calculated using data from the previous 12 months. 2016 year data was used only for past volatilities calculations.

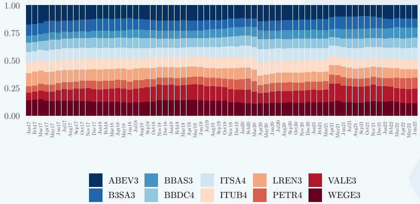


(a) Weigths

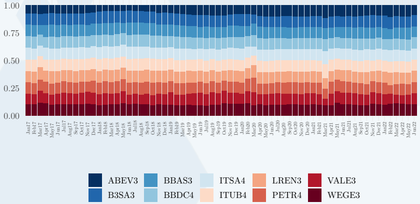


(b) Risks

Figure 7: Monthly distribution of MVP.



(a) Weigths



(b) Risks

Figure 8: Monthly distribution of RPP.

Figure 9: MVP and RPP accumulated returns: January 2017 – June 2022.

	MVP	RPP
Annualized Return	0.0785	0.1505
Annualized Std Dev	0.2419	0.2637
Annualized Sharpe (Rf=0%)	0.3246	0.5708

Table 1: Annualized returns: January 2017 - June 2022.

(a) January 2017 – December 2019.

(b) January 2020 – June 2022.

Figure 10: MVP and RPP accumulated returns.

	MVP	RPP
Annualized Return	0.2107	0.3212
Annualized Std Dev	0.1744	0.2027
Annualized Sharpe (Rf=0%)	1.2087	1.5851

Table 2: January 2017 - December 2019.

	MVP	RPP
Annualized Return	-0.0631	-0.0278
Annualized Std Dev	0.3046	0.3227
Annualized Sharpe (Rf=0%)	-0.2070	-0.0860

Table 3: January 2020 - June 2022.

Thanks for your attention!

References

[1] Ademir A Aguiar, Orizon P Ferreira, and Leandro F Prudente. “Inexact gradient projection method with relative error tolerance”. In: *arXiv preprint arXiv:2101.11146* (2021).

[2] Zeyuan Allen-Zhu et al. “Linear convergence of a Frank-Wolfe type algorithm over trace-norm balls”. In: *Advances in Neural Information Processing Systems*. 2017, pp. 6191–6200.

[3] Amir Beck and Marc Teboulle. “A conditional gradient method with linear rate of convergence for solving convex linear systems”. In: *Math. Methods Oper. Res.* 59.2 (2004), pp. 235–247. URL: https://doi.org/10.1007/s001860300327.

[4] Ernesto G. Birgin, José Mario Martínez, and Marcos Raydan. “Inexact spectral projected gradient methods on convex sets”. In: *IMA J. Numer. Anal.* 23.4 (2003), pp. 539–559.

[5] S. Bonettini, R. Zanella, and L. Zanni. “A scaled gradient projection method for constrained image deblurring”. In: *Inverse Problems* 25.1 (2009), pp. 015002, 23.

[6] James P. Boyle and Richard L. Dykstra. “A method for finding projections onto the intersection of convex sets in Hilbert spaces”. In: *Advances in order restricted statistical inference (Iowa City, Iowa, 1985)*. Vol. 37. Lect. Notes Stat. Springer, Berlin, 1986, pp. 28–47. URL: https://doi.org/10.1007/978-1-4613-9940-7_3.

[7] Richard L. Dykstra. “An algorithm for restricted least squares regression”. In: *J. Amer. Statist. Assoc.* 78.384 (1983), pp. 837–842. URL: http://links.jstor.org/sici?sici=0162-1459(198312)78:384%3C837:AAFRLS%3E2.0.CO;2-D&origin=MSN.

[8] D.S. Gonçalves, M.A. Gomes-Ruggiero, and C. Lavor. “A projected gradient method for optimization over density matrices”. In: *Optimization Methods and Software* 31.2 (2016), pp. 328–341.

[9] Geovani N. Grapiglia and Ekkehard W. Sachs. "On the worst-case evaluation complexity of non-monotone line search algorithms". In: *Comput. Optim. Appl.* 68.3 (2017), pp. 555–577.

[10] L. Grippo, F. Lampariello, and S. Lucidi. "A nonmonotone line search technique for Newton's method". In: *SIAM J. Numer. Anal.* 23.4 (1986), pp. 707–716.

[11] Martin Jaggi. "Revisiting Frank-Wolfe: Projection-free Sparse Convex Optimization". In: *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28* ICML'13 (2013), pp. I-427–I-435. URL: http://dl.acm.org/citation.cfm?id=3042817.3042867.

[12] H. Kaya and W. Lee. "Demystifying risk parity". In: *Neuberger Bermann* (2012).

[13] Harry Markowitz. "Portfolio Selection". In: *The Journal of Finance* 7.1 (1952), pp. 77–91.

[14] Fabiana R. de Oliveira, Orizon P. Ferreira, and Gilson N. Silva. “Newton’s method with feasible inexact projections for solving constrained generalized equations”. In: *Comput. Optim. Appl.* 72.1 (2019), pp. 159–177.

[15] Marcos Raydan and Pablo Tarazaga. “Primal and polar approach for computing the symmetric diagonally dominant projection”. In: *Numerical Linear Algebra with Applications* 9.5 (2002), pp. 333–345. DOI: https://doi.org/10.1002/nla.277. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nla.277. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/nla.277.

[16] Saverio Salzo and Silvia Villa. “Inexact and accelerated proximal point algorithms”. In: *J. Convex Anal.* 19.4 (2012), pp. 1167–1192.

[17] H. Zhang and W. W. Hager. “A nonmonotone line search technique and its application to unconstrained optimization”. In: *SIAM J. Optim.* 14.4 (2004), pp. 1043–1056.

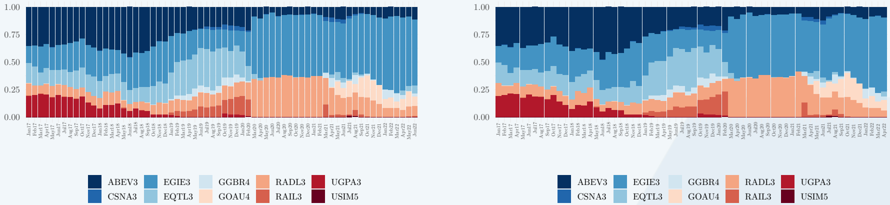
Backtest Study - 2

We present an application for Algorithm InexProj-SGM, a backtest study of a portfolio build with the 5 most volatile brazilian companies of 2016: Gerdau S.A. (GGBR4, GOAU4), CSN S.A (CSNA3), RUMO S.A. (RAIL3), Usiminas (USIM5) and the 5 least volatile assets of 2016: Ambev S.A. (ABEV3), Engie S.A. (EGIE3), Ultrapar S.A. (UGPA3), Equatorial (EQTL3), Raia Drogasil S.A. (RADL3).

We did a backtest, building 2 portfolios with monthly rebalance:

- One was the minimum variance portfolio (MVP),
- the other the risk parity portfolio (RPP).

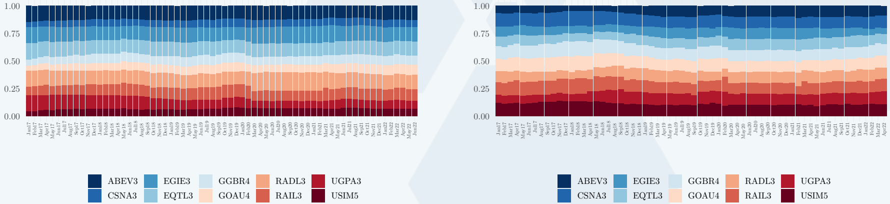
Data pertaining to the period between January 2016 and June 2022, was taken from Yahoo Finance (`http://finance.yahoo.com`). For rebalancing, the volatilities were calculated using data from the previous 12 months. 2016 year data was used only for past volatilities calculations.



(a) Weigths

(b) Risks

Figure 11: Monthly distribution of MVP.



(a) Weigths

(b) Risks

Figure 12: Monthly distribution of RPP.

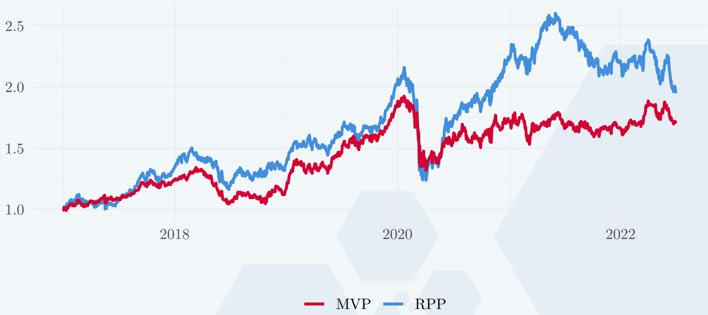
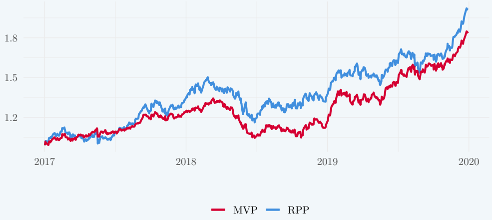


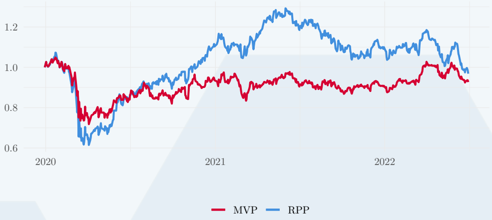
Figure 13: MVP and RPP accumulated returns: January 2017 – June 2022.

	MVP	RPP
Annualized Return	0.1035	0.1317
Annualized Std Dev	0.2122	0.2552
Annualized Sharpe (Rf=0%)	0.4880	0.5160

Table 4: Annualized Returns: Jan17 - Jun22



(a) January 2017 — December 2019.



(b) January 2020 — June 2022.

Figure 14: MVP and RPP accumulated returns.

	MVP	RPP
Annualized Return	0.2298	0.2688
Annualized Std Dev	0.1590	0.1943
Annualized Sharpe (Rf=0%)	1.4447	1.3836

Table 5: Annualized Returns: Jan17 - Dez19

	MVP	RPP
Annualized Return	-0.0305	-0.0130
Annualized Std Dev	0.2619	0.3128
Annualized Sharpe (Rf=0%)	-0.1165	-0.0415

Table 6: Annualized Returns: Jan20 - Jun22